METHOD AND SYSTEM FOR QUALITY OF SERVICE (QoS) SUPPORT IN A PACKET-SWITCHED NETWORK

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/202,190, entitled INTERNET PROTOCOL TRANSPORT, filed May 5, 2000 which is hereby incorporated by reference.

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to the field of telecommunication networks, and more particularly to a method and system for quality of service (QoS) support in a packet-switched network.

BACKGROUND OF THE INVENTION

Telecommunication networks transport voice and data according to a variety of standards and using a variety of technologies. Circuit-switch networks such as plain old telephone service (POTS) utilize transmission paths dedicated to specific users for the duration of a call and employ continuous, fixed-bandwidth transmission. Packet-switch networks (PSNs) allow dynamic bandwidth, depending on the application, and can be divided into connectionless networks with no dedicated paths and connection-oriented networks with virtual circuits having dedicated bandwidth along a predetermined path. Because packet-switched networks allow traffic from multiple users to share communication links, these networks utilize available bandwidth more efficiently than circuit-switched networks.

Internet protocol (IP) networks are connectionless packet-switched networks. IP networks transport information by breaking up bitstreams into addressable digital packets. Each IP packet includes source and destination addresses and can take any available route between the source and the destination. The IP packets are transmitted independently and then reassembled in the correct sequence at the destination.

Conventional IP networks employ routers to direct packets to their destination. Packets are inspected at each router for network protocol addresses and forwarded to a next router on the way toward the destination based on downstream congestion and other real-time parameters. While this inspection and dynamic routing provides a high degree of flexibility within the network, it adds delays to each packet at each router. Accordingly, traffic

transported across an IP network between geographically distributed source and destination points will have a relatively large cumulative delay. This limits the ability of the IP network to support voice, video, and other real-time applications.

SUMMARY OF THE INVENTION

The present invention provides an improved method and system for transporting traffic in a packet-switched network that substantially eliminate or reduce the problems and disadvantages associated with previous systems and methods. In a particular embodiment, the present invention maps external quality of service (QoS) classes into a reduced set of internally defined QoS classes while supporting essential features of the external QoS classes.

In accordance with one embodiment of the present invention, a method and system for transporting traffic having disparate qualities of service classes across a packet-switched network includes receiving at an ingress node of a private or other suitable network a plurality of packets each having a quality of service (QoS) class defined externally to the network. Packets having a QoS class including delay bound guarantees and a low drop priority are combined into a first internal QoS class. Packets having a QoS class including a flexible drop priority and no delay bound guarantees are combined into a second internal QoS class. Packets having a QoS class including no delivery guarantees are combined into a third internal QoS class. The packets are transmitted across the network based on their internal QoS classes.

Technical advantages of the present invention include providing an improved packet-switched network. In a particular embodiment, QoS support is efficiently provided for the packet-switched network. This allows provisioning of enhanced services and service differentiation by the network provider without high implementation cost.

Another technical advantage of the present invention includes providing voice, video and other real-time support for Internet protocol (IP) traffic using a partial QoS feature set in which non-essential features of well defined or standardized QoSs are combined into behavior aggregate classes. As a result, the number of internal QoS classes is reduced and the complexity to network reduced.

Other technical advantages of the present invention will be readily apparent to one skilled in the art from the following figures, description, and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention and its advantages, reference is now made to the following description taken in conjunction with the accompanying drawings, wherein like reference numerals represent like parts, in which:

FIGURE 1 is a block diagram illustrating a transport network in accordance with one embodiment of the present invention;

FIGURE 2 is a block diagram illustrating an external representation for the transport router of FIGURE 1 in accordance with one embodiment of the present invention;

FIGURE 3 is a block diagram illustrating details of the Internet protocol transport (IPT) node of FIGURE 1 in accordance with one embodiment of the present invention;

FIGURE 4 is a block diagram illustrating a fast transport segment (FTS) defined through the transport network of FIGURE 1 in accordance with one embodiment of the present invention;

FIGURE 5 is a block diagram illustrating details of the receiver-transmitter pair (RTP) of FIGURE 3 in accordance with one embodiment of the present invention;

FIGURE 6 is a block diagram illustrating combining defined quality of service (QoS) classes into IPT QoS classes for transport in the network of FIGURE 1 in accordance with one embodiment of the present invention;

FIGURE 7 is a block diagram illustrating traffic flow through the RTP of FIGURE 5 in accordance with one embodiment of the present invention; and

FIGURE 8 is a flow diagram illustrating a method for processing traffic for QoS-based transport through the

transport network of FIGURE 1 in accordance with one embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

FIGURE 1 illustrates a transport network 10 in accordance with one embodiment of the present invention. In this embodiment, the transport network 10 is an Internet protocol (IP) network for transporting IP and Multiple Protocol Label Switch (MPLS) packets. The transport network 10 may be any other packet-switched network operable to route, switch, and/or otherwise direct data packets based on network protocol addresses.

The transport network 10 is a private network connecting geographically distributed segments of an external network 12. The external network 12 includes one or more public and/or private networks such as the Internet, an intranet, and other suitable local area networks (LAN), wide area networks (WAN), and nodes. The external network 12 includes label switch and subtending routers 14, Ethernet switches 16, Frame Relay switches 18 and other suitable routers, switches, and nodes operable to generate and/or transport traffic. The transport network 10 communicates with nodes of the external network 12 in the native protocol of the nodes to communicate traffic and control signaling between the networks 10 and 12.

Referring to FIGURE 1, the transport network 10 includes a plurality of Internet protocol transport (IPT) nodes 30 interconnected by communication links 32. The IPT nodes 30 each include a plurality of ports 34 accessible to the external network 12. As used herein, each means every one of at least a subset of the identified items. The communication links 32 are optical fiber or other suitable high-speed links. As described

in more detail below, the high-speed links 32 connect high speed interfaces of the IPT nodes 30 to form fast transport segments (FTS) through the transport network 10. Packets transferred via the FTSs incur very small buffering delay in the network. Packets carried through the ports 34 and between FTSs may incur queuing delay comparable to a normal IP switch.

To optimize bandwidth usage within the transport network 10, packets may be transmitted directly on the high-speed optical links 32 without synchronous optical network (SONET) framing and its associated overhead which imposes a penalty of three to five percent depending on the line rate. In one embodiment, a transport label is added to each packet to generate an internal packet that can be directly transmitted on the optical links 32. Transport label may be include format information indicating the type of signal being transmitted, a label value including a destination network address for a connectionless flow or a path identifier for a connection-oriented flow, a quality of service (QoS) identifier, an end-of-stack indicator, and time-to-live information. Details of the transport label are described in co-owned U.S. Patent Application entitled *"System and Method for Connectionless/Connection Oriented Signal Transport"*, filed June 6, 2000. Using the transport label, both connection-oriented and connectionless traffic may be seamlessly transported across the transport network 10. Protection for connection oriented data flows may be provided as described in co-owned U.S. Patent Application entitled *"Method and System For Providing A Protection Path For Connection-Oriented Signals In A Telecommunications*

Network", filed June 6, 2000. Protection for connectionless traffic flows may be provided as described in co-owned U.S. Patent Application "*Method and System For Providing A Protection Path For Connectionless Signals In A Telecommunications Network*", filed June 6, 2000.

To support voice, video, and other real-time or time-sensitive applications, the transport network 10 provides quality of service (QoS), which may include class of service (CoS), differentiation. In one embodiment, all IP packets are mapped to one of three priority levels as they enter the transport network 10. In this embodiment, guaranteed traffic has reserved bandwidth and is guaranteed to be transported within a defined time delay. Control flow traffic is also reserved and guaranteed, but the network 10 does not guarantee delivery time delay. Best effort traffic does not have reserved bandwidth and delivery is not guaranteed by the network 10. By distinguishing and prioritizing traffic based on its QoS priority, including CoS and/or service level agreement (SLA), and/or other suitable indication of importance or delivery constraints, the transport network 10 is able to deliver time-sensitive traffic within tight time constraints by delaying and/or dropping best effort traffic and other low priority traffic.

In one embodiment, the transport network 10 utilizes a private internal addressing scheme to isolate the network 10 from customers and thus minimize or prevent conflicts with private and/or public networks connected to the transport network 10. This reduces the complexity of network management and preserves the topology of the

existing routed network 12. In addition, transport network isolation enables value added services to be provided through the transport network 10.

When an independent addressing scheme is utilized for the transport network 10, egress traffic is converted from the external addressing scheme to the internal addressing scheme at ports 34 using standardized or extended network address translation (NAT). Similarly, egress traffic is converted from the internal addressing scheme back to the external addressing scheme at ports 34 using standard or extended NAT. In addition to the internal addresses, each IPT node 30, port 34 and other component of the transport network 10 visible to the external network 12 includes a globally unique IP address. These addresses are used for external management of the transport network 10.

The transport network 10 provides a flexible topology in which sets of ports 34 may be grouped in any suitable way and each treated as a single entity capable of independently interacting with external nodes. Thus, the transport network 10 is externally represented as sets of port groups 50 with internally managed connectivity. Provisioning of port groups 50 in the transport network 10 is unconstrained with mesh and partial-mesh topologies supported.

The port groups 50 are each a set of ports 34 with similar routing properties. In particular, a port group 50 is a set of ports 34 configured to provide multipoint-to-multipoint or at least point-to-multipoint connectivity between each other which allows point-to-multipoint connectivity between external elements. Accordingly, traffic received by a port group 50 can be

routed directly from an ingress port 34 to a plurality of egress ports 34 without channelization in the transport network 10.

Port groups 50 may be provisioned as simple port groups and as composite port groups. In the simple port group configuration, each port 34 only belongs to a single port group 50. Private addresses can be supported inside the simple port group configuration. A composite port group includes ports 34 which have membership in multiple port groups 50. In the composite port group case, private IP addressing is not supported.

The port groups 50 each define a transport element 52 with geographically distributed ports 34. Each transport element 52 is assigned a unique global IP address for peering and protocol exchanges within and/or external to the transport network 10. As described in more detail below, the transport elements 52 may implement a distributed architecture in which local processors control each of the ports 34 and a centralized processor controls the network element 52.

In particular embodiments, the transport elements may be transport routers 60 interconnecting sets of subtending IP routers 14, transport Ethernet switches 62 interconnecting sets of subtending Ethernet switches 16, and transport Frame Relay switches 64 interconnecting sets of subtending Frame Relay switches 18. In addition, the transport element 52 may interconnect two ports transparently, in which case the port group 50 is user protocol independent.

FIGURE 2 illustrates details of the transport router 60 in accordance with one embodiment of the present invention. In this embodiment, the transport router 60

comprises a simple port group and acts as a single network element within a customer's autonomous network.

Referring to FIGURE 2, the transport router 60 includes geographically distributed ports 34 connected to external routers 14. The external ports 34 form a port group 50 with point-to-multipoint connectivity between the ports 34 as externally represented by the router 80. Accordingly, traffic from any one of the external routers 14 may be routed from an ingress port 34 directly to any number of the other external routers 14 by router 80.

The transport router 60 includes a router identifier to peer with the external routers 14 and participate in reservation and other protocol exchanges. In a particular embodiment, the transport router 60 peers with subtending routers 14 by using interior gateway protocols (IGP) such as OSPF, IS-IS, or RIP. The transport router 60 may peer using an exterior gateway protocol (EGP) or any other suitable protocol.

FIGURE 3 illustrates details of the IPT node 30 in accordance with one embodiment of the present invention. In this embodiment, the IPT node 30 comprises an add/drop multiplexer (ADM) with modular building blocks to support a scalable, pay-as-you-grow architecture. Accordingly, the transport network 10 owner may add functionality and incur cost based on customer demand. The IPT node 30 comprises logic encoded in software or hardware media for performing functions of the node. The logic may be distributed between discrete cards in the node.

Referring to FIGURE 3, the IPT node 30 includes one or more receiver-transceiver pairs (RTP) 100 and a processing system 102 interconnected by an internal Ethernet connection. As described in more detail below,

each RTP 100 includes one or more internal interfaces 104 and one or more local interfaces 106. The internal interfaces are high-speed interfaces between the IPT nodes 30 while the local interfaces 106 are low-speed ports 34 accessible to external nodes and/or interfaces between FTSs.

Within the transport network 10, a set of internal interfaces 104 of the IPT nodes 30 are connected together between ports 34 of a port group 50 to form an FTS between the ports 34 and provide multipoint-to-multipoint and/or point-to-multipoint connectivity. In particular, a multiplexer of an internal interface 104 is connected to a demultiplexer of a next internal interface 104 in the FTS while a demultiplexer of the internal interface 104 is connected to a multiplexer of a previous internal interface 104 in the FTS. The FTSs are directionally-sensitive to preferentially route pass-through traffic over local ingress traffic. In this way, traffic for a transport element 52 is transported between an ingress and an egress port on an FTS to minimize delay within the transport network 10.

The processing system 102 includes one or more central processing units (CPUs) 108. The CPUs 108 may each operate the IPT node 30 or a transport element 52. A CPU 108 operating the IPT node 30 includes an operating system and control functionality for the IPT node 30. A CPU 108 operating a transport element 52 includes control functionality for the distributed components of the transport element 52.

FIGURE 4 illustrates a FTS 110 in accordance with one embodiment of the present invention. In this embodiment, the FTS 110 comprises 10 Gb/s links and

directionally-sensitive interfaces to provide a cumulative delay of less than 2.5 microseconds for a 1,500 bite maximum packet size. It will be understood that the FTS 110 may comprise other high-speed links and interfaces. A high-speed link is operable to transport traffic at a rate of 5 Gb/s or greater. Preferably, the high-speed links transport traffic at rates of 10 Gb/s or above.

Referring to FIGURE 4, the FTS 110 comprises dedicated internal interfaces 104 and high-speed links 32 extending from a source node 112 through a plurality of intermediate nodes 114 to a destination node 116. A local interface 106 is coupled to each of the internal interfaces 104 to allow local traffic to be added and dropped from the FTS 110.

As described in more detail below, in the FTS 110, each internal interface 104 segments local and pass-through traffic. The local traffic is dropped. The pass-through traffic is segmented into high and low priority pass-through traffic. The high priority pass-through traffic is transmitted along the FTS 110 preferentially over the low priority pass-through traffic and local ingress traffic from the local interface 106. The low priority pass-through is buffered. A traffic class is transmitted preferentially over other traffic when it is transferred first using needed bandwidths, the other traffic using remaining bandwidth for transmission.

The local traffic is segmented into high priority local traffic and low priority local traffic. The high priority local traffic is transmitted preferentially over the low priority pass-through traffic and the low priority local traffic. Accordingly, high priority pass-

through traffic is transmitted without or with only minimum delay while avoiding starvation at the intermediate nodes 114.

The low priority traffic is transmitted based on remaining bandwidth availability. In one embodiment, the low priority pass-through traffic is transmitted preferentially over the low priority local traffic to give preference to pass-through traffic at all priorities. The high priority traffic may be reserve bandwidth traffic and the low priority traffic unreserved bandwidth traffic. In a particular embodiment, as described in more detail below, the high-priority traffic comprises internally defined guaranteed service and control load traffic and the low-priority traffic comprises internally defined best-effort traffic. Additional and intermediate priorities of traffic may be identified, segmented, and used to preferentially route traffic in the network.

In a particular embodiment, local and pass-through traffic is distinguished and segmented based on a shallow IP layer 2/3 lookup using the transport label. In this embodiment, the transport label identifies the corresponding packet as local or remote (pass-through) and identifies the internal QoS of the packet. Local traffic is dropped while the priority of the pass-through traffic is determined based on QoS for immediate transmission out or buffering. Similarly, ingress local traffic is labeled and analyzed to determine its transmission priority. Traffic having the same priority is transmitted in a first-in/first-out (FIFO) basis.

FIGURE 5 illustrates details of the RTP 100 in accordance with one embodiment of the present invention.

In this embodiment, the internal interface 104 is a high-speed interface that operates at substantially 10 Gb/s. The external interface 106 is a low-speed packet over SONET (POS) interface that operates at 2.5 Gb/s or below.

Referring to FIGURE 5, the internal interface 104 includes an optical receiver 120, a demultiplexer 122, a multiplexer 124, and an optical transmitter 126. The optical receiver is a 10 Gb/s receiver without SONET or package level knowledge. The optical receiver 120 performs the optical to electrical signal conversion. The optical receiver 120 may include an amplifier and may directly interface with a wave division multiplex (WDM) system.

The demultiplexer 122 drops local traffic and inter RTP traffic as well as buffers transit traffic. In a particular embodiment, the demultiplexer 122 has a set of 155 Mb/s connections to interface cards of the external interface 106. The demultiplexer 122 may also have 155 Mb/s connections to interface cards of other RTPs 100.

The multiplexer 124 collects local traffic from the interface cards of the external interface 106 and through traffic from the demultiplexer 122. The multiplexer 124 includes packet buffer, scheduler and insertion control functionality.

The optical transmitter 126 is a 10 Gb/s transmitter without SONET or package level knowledge. The optical transmitter 126 may include an optical amplifier. The optical transmitter 126 performs a conversion from an electrical signal to an optical signal and may interface directly with a WDM system.

The local interface 106 include a plurality of low-speed interface cards 130. The low-speed interface cards

130 send and receive traffic to and from the multiplexer 124 and demultiplexer 122, respectively. The low-speed interface cards 130 also provide connections between the FTSs.

The low-speed interface cards 130 are the main buffering point for ingress and egress traffic of the transport network 10. Packet level intelligence, including routing and protection mechanisms, are provided by the low-speed interface cards 130. If the transport network 10 uses an isolated addressing scheme, the low-speed interface cards 130 perform NAT functionality.

In a particular embodiment, low-speed interface cards 130 each include a buffer for each internal QoS class, which as previously described, may be guaranteed service, control load and best effort. In this and other embodiments, each buffer may discard packets based on its own thresholds, independent of the others. Because guaranteed service and control-load traffic have reserved paths, comforming traffic typically will not be dropped. Best-effort traffic will be dropped based on congestion at the node.

Traffic received by the interface cards 130 from external links are associated with a corresponding data flow and a transport label generated for and/or added to packet for transport through the network. In generating the label, the interface card 130 maps the external QoS class to one of the reduced number of internal QoS classes. The external QoS classes are defined outside or independently of the transport, private or other suitable network and may be well-defined classes such as standardized classes. The internal QoS classes are defined by and/or within the network. The packet with

the appended label is queued in a corresponding buffer and transmitted across the network along a path identified by the label and based on its internal QoS class. To provide a QoS guarantee for each new traffic flow, a path through the network that has sufficient resources to meet the flow's requirements is identified. The flow's requirements may be bandwidth and/or delay guarantees. In one embodiment, feasible paths are dynamically determined based on availability of network resources throughout the network. In this embodiment, network resources are stored in a link state database in the IPT nodes 30 which are provisioned and/or updated using opaque link state advertisement (LSA) to advertise the available link bandwidth and propagation delay in the network.

A constraint shortest path first (CSPF), open shortest path first (OSPF) or the suitable algorithm may utilize the link state database to compute feasible paths and/or optimal paths. The CSPF and/or OSPF algorithms may comprise Bellman-Ford or Dijkstra algorithms and may be optimized for one cost, such as bandwidth or propagation delay. To satisfy both requirements for a connection, sequential filtering may be used. In this embodiment, paths based on bandwidth or other primary metric are computed first and a subset of them eliminated based on propagation delay or other secondary metric until a single, optimum or acceptable path is found. For a guaranteed service or control load traffic, bandwidth for the path is reserved using a signaling or other suitable protocol. Paths may be reserved as described in co-owned US patent application entitled "*System and Method for Application Object Transport*", filed June 6,

2000. It will be understood that suitable, preferred and/or optimal paths may be otherwise determined based on availability of network resources and that bandwidth may be otherwise suitably reserved in the transport network. For example, paths may be identified through the transport network by pruning links that do no have enough bandwidth to meet the bandwidth request with a probability greater than a defined value and minimum-hop paths computed based on the pruned topology. In this embodiment, if there are two or more minimum-hop paths, the one with the largest available bandwidth may be chosen. In addition, to account for inaccurate information on network resource availability, a weighing metric may be used and adjusted to account for the probability distribution function. In a particular embodiment, OSPF and IS-IS extensions carried via opaque LSA's may be used to gather resource availability. In a particular embodiment, the extension comprises Router Address TLV (type 1), Link TLV (type 2), Link Type sub-TLV (sub-type 1), Link ID sub-TLV (sub-type 2), Local Interface IP Address sub-TLV (sub-type 3), Remote Interface IP Address sub-TLV (sub-type 4), Traffic Engineering Metric sub-TLV (sub-type 5), Maximum Bandwidth sub-TLV (sub-type 6), Maximum Reservable Bandwidth sub-TLV (sub-type 7), Unreserved Bandwidth sub-TLV (sub-type 8), Resource Class/Color sub-TLV (sub-type 9), Router ID TLV (type 134), Extended IP Reachability TLV (type 135), Extended IS Reachability TLV (type 22), Administrative Group sub-TLV (sub-type 3), IPV4 Interface Address sub-TLV (sub-type 6), IPv4 Neighbour Address sub-TLV (sub-type 8), Maximum Link Bandwidth sub-TLV (sub-type 9), Maximum Reservable Link Bandwidth sub-TLV (sub-

type 10), Unreserved Bandwidth sub-TLV (sub-type 11), TE Default Metric sub-TLV (sub-type 18).

FIGURE 6 illustrates combining packets having disparate external QoS classes into internal IPT QoS classes for transport in the network 10 in accordance with one embodiment of the present invention. In this embodiment, integrated and differentiated services classes are combined into three internal IPT classes. It will be understood that other standardized or well-defined service classes may be similarly combined into the reduced or other suitable set of internal QoS classes without departing from the scope of the present invention.

Referring to FIGURE 6, the internal QoS classes include an IPT guaranteed service class (gs) 140, an IPT control load (CL) class 142, and an IPT best effort (BE) class 144. In an particular embodiment, the IPT GS class 140 is characterized by low latency with delayed bound guarantees and a low drop priority. This service utilizes reservation. The IPT CL class 142 is characterized with no delay bound guarantees but with flexible drop priority. This class also uses reservation. The IPT BE class 144 provides no delivery guarantees in accordance with transmission of standard data traffic over the Internet. The IPT classes 140, 142 and 144 together support a subset of the standardized QoS features with non-essential features combined to reduce the number of QoS classes, which may reduce the cost and complexity of the network 10. Accordingly, the IPT classes 140, 142 and 144 each represent a queuing behavior type and/or behavior aggregate.

For the standardized integrated services QoS classes, the guaranteed class 146 is mapped into the IPT GS class 148. Guaranteed class 146 provides an assured level of bandwidth that when used by a policed flow produces a delay bounded service with no queuing loss for all comforming packets. The guaranteed service 146 does not attempt to control the minimal or average delay of a packet, but controls the maximum queuing delay. The guaranteed service 146 guarantees the packets will arrive within the guaranteed delivery time and will not be discarded due to queued overflows, provided that flow's traffic stays within the specified traffic parameters. The service is used by real-time applications that require packets to arrive no later than a certain time after transmission by a source. For example, the guaranteed service 146 may be used by audio and video play-back applications.

The integrated services control load class 148 is mapped into the IPT CL class 142. The control load class 148 provides an end-to-end behavior tightly approximating the behavior visible to applications receiving best-effort service under unloaded conditions. A very high percentage of transmitted packets will be successfully delivered by the network to the receiving end nodes. The transit delay experience by a high percentage of the delivered packets will not greatly exceed the minimum transmit delay experience by any successfully delivered packet. Should traffic fall outside an estimated amount, a large number of packets may be delayed or dropped.

For the differentiated services, the expedited forwarding class 150 provides a low loss, low latency, low jitter, assured bandwidth end-to-end service. The

service appears to the endpoints like a point-to-point connection by a virtual lease line. The expedited forwarding class 150 is mapped to the IPT GS class 140 along with the guaranteed class 146. Accordingly, the IPT GS class comprises traffic within assured level of bandwidth, and low loss, low latency, low jitter or other delay-bounded requirements corresponding to standardized classes. In one embodiment, each metric for defining an internal class 140, 142 and/or 144 may be the strictest metric of the combined standardized classes.

The differentiated services assured forwarding classes (1, 2 and 3) 152 provide separate levels of forwarding assurances with one of three possible drop-precedence values. In case of congestion, the drop precedence of a packet determines the relative importance of a packet within the assured forwarding class 152. Packets with a lower-drop precedence value are protected by preferably discarding packets with a higher drop precedence value. Each assured forwarding group is provided a minimum forwarding bandwidth assurance, and any excessive bandwidth is fairly shared. The assured forwarding groups 1, 2 and 3 are mapped to the IPT CL class 142 along with the control load services 148. Accordingly, the IPT CL class comprises traffic with no delay bound and a flexible drop priority in accordance with the corresponding defined service classes. The IPT CL Class 142 has no specified latency but may require reservation by signaling or SLA.

Differentiated services assured forwarding 4 class 154 is mapped into the IPT BE class 144 along with the differentiated services best effort class 156. Alternatively, assured forwarding group 4 may be

supported by the IPT CL class 154. The IPT BE class provides no latency limits or reservation. In this way, a large number of externally-defined and/or standardized QoS classes can be supported by a reduced set of internally defined QoS classes, which support the most important features of the defined external QoS classes.

FIGURE 7 illustrates traffic flows through the RTP 100 in accordance with one embodiment of the present invention. In this embodiment, traffic is distinguished, segregated, and processed based on a two level, low/high priority scheme, with the IPT GS and IPT CL classes 140 and 142 comprising the high priority traffic. It will be understood that the traffic flows may be segmented into any number of suitable traffic types based on QoS and other suitable traffic type identifiers.

Referring to FIGURE 7, the RTP 100 includes internal interface 104 and local interface 106. The internal interface 104 includes the receiver 120, demultiplexer 122, multiplexer 124 and transmitter 126. A traffic buffer 150 is coupled between the demultiplexer 122 and multiplexer 124. The local interface 106 includes a local buffer 152 coupled between the demultiplexer 122, multiplexer 124 and a local port 154.

The receiver 120 includes an optical to electrical interface (OEI) 160 for converting ingress optical signals from the high-speed optical links 32 to electrical signals. The demultiplexer 122 includes a lookup table 162 for identifying pass-through and local traffic. The transmitter 126 includes an OEI 164 for converting an egress traffic stream to optical signals for transmission over the high-speed optical links 32. The transmit buffer 150 is a two packet or other suitable

sized buffer operable to hold direct pass-through packets while the multiplexer 124 completes processing of a current packet.

The local buffer 152 receives low priority pass-through traffic of the IPT BE class 144 and buffers the traffic for transmission based on bandwidth availability. Egress local traffic is dropped through the local buffer 152 to the local port 154 for transmission to a local designation or another FTS 110. The local buffer 152 also receives and buffers ingress high and low priority local traffic for transmission on the FTS 110 based on bandwidth availability. Local buffer 152 may include a scheduler 170 to shape low priority pass-through and local traffic.

The local port 152 receives and transmits local traffic. In one embodiment, the local port 152 includes a demultiplexer 172 with lookup table 174 for distinguishing and segmenting high and low priority ingress local traffic. This allows all high priority traffic to be transmitted preferentially over all low priority traffic regardless of the source and/or the destination of the traffic. Within the high priority traffic, packets from the IPT GS class 140 may be preferentially transmitted over packets in the IPT CL class 142.

In operation, an ingress traffic stream is received at the receiver 120 and converted to an electrical packet stream 178 by OEI 160. The packet stream 178 is demultiplexed by demultiplexer 122 into discrete packets and segmented into local egress traffic 180 and pass-through traffic 182. The pass-through traffic 182 is further segmented based on its QoS into high priority

pass-through traffic 184 and low priority pass-through traffic 186.

The high priority pass-through traffic 186 is passed to the multiplexer 124 through the transmit buffer 150 while the low priority pass-through traffic 186 is dropped to the local buffer 152. The local buffer 152 drops egress local traffic 180 and hairpins low priority pass-through traffic 186 for transmission back on the FTS 110 based on bandwidth availability.

Local ingress traffic is demultiplexed at the local port 154 and segmented into high priority local ingress traffic 190 and low priority local ingress traffic 192 using the lookup table 174. The local buffer 154 receives and buffers the high and low priority local traffic 190 and 192 along with the low-priority pass-through traffic 186.

The multiplexer 124 inserts all high-priority pass-through traffic from the transmit buffer 150 into an egress traffic flow 194 immediately or, if active, immediately upon finishing a current packet. High priority local traffic 190 is inserted into available bandwidth with the low priority pass-through local traffic inserted into the remaining available bandwidth of the egress traffic flow and with the low priority pass-through and local traffic inserted into the remaining available bandwidth. The multiplexer 124 multiplexes the traffic flows into an egress traffic stream 194 that is converted to an optical signal by OEI 164 for transmission over the high speed optical link 32. In this way, high priority pass-through traffic passes the RTP 100 with little or no delay. Local high priority traffic is delayed transmission on the FTS 110 until

bandwidth first becomes available. After that point, it is treated as pass-through traffic by downstream nodes to prevent additional delays. Accordingly, queuing delays can be estimated and are minimized in the network, which increases bandwidth manageability and applications that can be supported by the network.

FIGURE 8 is a flow diagram illustrating a method for processing traffic in a node for QoS-based transport across the transport network 10. The method begins at step 250 in which an ingress traffic stream is received. At step 252, the ingress traffic stream is demultiplexed into individual IP packets.

Proceeding to step 254, local and pass-through traffic is segregated. The traffic may be segregated using the transport label and a shallow lookup or a standard routing table lookup. At step 256, high and low priority pass-through traffic is segregated. In one embodiment, guaranteed, control load and control signal traffic using reserve bandwidth are treated as high priority traffic while best effort traffic using unreserved bandwidth is treated as low priority traffic.

Proceeding to step 258, local egress traffic is dropped. At step 260, low priority pass-through traffic is buffered. At step 266, high priority pass-through traffic is inserted into the egress flow for immediate transmission regardless of the amount of local traffic waiting transmission. At step 268, high priority local traffic is inserted into the egress flow based on bandwidth availability with guaranteed traffic given preference. Thus, all high-priority traffic is transmitted before low priority traffic is processed

regardless of the source or destination of the low priority traffic.

Next, at step 270, the low priority traffic is inserted into the egress traffic flow based on remaining bandwidth availability. The low priority traffic may be inserted in a FIFO order or preferentially with pass-through traffic transmitted prior to local traffic. At step 272, the egress flows are multiplexed into an egress traffic stream. The egress traffic stream is transmitted on the FTS 110 at step 274. In this way, substantial cumulative delays are avoided in the network.

Although the present invention has been described with several embodiments, various changes and modifications may be suggested to one skilled in the art. It is intended that the present invention encompass such changes and modifications as fall within the scope of the appended claims.